

Construção de escalas de proficiência para exames seletivos, por meio da Teoria da Resposta ao Item

Edilan de Sant'Ana Quaresma ¹

Pedro Alberto Barbeta ²

Adriano Ferreti Borgatto ²

Ezequiel Abraham López Bautista ³

Carlos Tadeu dos Santos Dias ^{4 5}

1 Introdução

Os processos seletivos vestibulares passaram a ser a forma padrão de ingresso nas universidades brasileiras, a partir dos anos 70 do século XX. Historicamente estes processos de avaliação utilizaram a Teoria Clássica dos Testes na classificação dos candidatos, sem considerar as informações contidas em cada um dos itens que compõe o exame seletivo, fazendo com que o conhecimento dos candidatos fosse medido unicamente de acordo com o número de questões respondidas corretamente. Thrustone (1928) questionava os testes como instrumento para medir o conhecimento dos alunos, ao considerar que conhecimento é uma variável latente, não sendo possível sua medida de maneira direta.

Nesse contexto emergiu a necessidade de uma teoria que viabilizasse a mensuração de variáveis latentes, aqui entendidas como habilidades ou proficiências de sujeitos, considerando cada um dos itens de uma prova individualmente e não somente o número de questões respondidas corretamente pelos candidatos. Uma ferramenta muito útil na resolução desse problema, começou a ser formulada nos anos 50 do século XX, e atualmente é denominada Teoria da Resposta ao Item - TRI. Vem sendo utilizada com sucesso no exame nacional de ensino médio - ENEM, cujos resultados têm sido adotados por universidades como critério de ingresso nos seus quadros discentes.

O trabalho aqui apresentado objetiva utilizar a TRI no padrão de resposta dos candidatos aos itens do exame vestibular coordenado pela FUVEST, nos seis cursos de graduação da Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, no ano de 2012, considerando somente a primeira fase do processo seletivo, para construção de uma escala de proficiências que melhor avalie os sujeitos.

¹ICED - UFOPA, LCE - ESALQ/USP. e-mail: equaresm@ufpa.br

²CTC - UFSC.

³Facultad de Agronomía, Universidad de San Carlos de Guatemala, LCE - ESALQ/USP

⁴LCE - ESALQ/USP

⁵Agradecimento ao CNPq pelo apoio financeiro.

2 Material e métodos

Os dados utilizados na pesquisa aqui proposta são provenientes do exame vestibular coordenado pela FUVEST e realizado no ano 2012, para os 2326 candidatos a ingressar em um dos seis cursos de graduação da ESALQ/USP.

Uma análise exploratória foi feita, utilizando o pacote LTM do R. Como cada uma das 89 questões válidas foi constituída por itens dicotômicos, em que as respostas dos candidatos aos itens foram do tipo certo ou errado, o modelo da TRI proposto para ser utilizado foi o de três parâmetros, definido como:

$$P(U_{ij} = 1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-Da_i(\theta_j - b_i)}}, \quad (1)$$

com $i=1,2,\dots,I$, e $j=1,2,\dots,n$, com parâmetros descritos em Andrade, Tavares e Valle (2000).

O modelo (1) foi ajustado ao padrão de respostas dos candidatos, por meio do software BILOG de tal forma que as estimativas dos parâmetros do modelo foram utilizadas como critério de seleção dos itens considerados importantes no processo de estimação das proficiências dos candidatos. Um critério usual para eliminar itens, avaliando unicamente o parâmetro “ a ” sugere considerar: (i) Itens cujo valor estimado para o parâmetro “ a ” sejam menores que 0,7; (ii) Itens cujo valor estimado para o parâmetro “ a ” apresentem um erro padrão muito grande, em relação aos demais itens padrões com o parâmetro “ a ” na mesma magnitude.

Uma grande vantagem da TRI no processo de avaliação dos itens, é a possibilidade de construção de uma escala de habilidades que possa ser pedagogicamente interpretável e, dessa forma, contribua para a classificação de sujeitos. Este processo de construção de uma escala de proficiências é definido por níveis âncora, caracterizados por conjuntos de itens chamados de itens âncora (ANDRADE, TAVARES e VALLE, 2000). Níveis âncora são pontos selecionados pelo analista na escala de proficiências para serem interpretados pedagogicamente, ao passo que itens âncora são itens selecionados.

2.1 Resultados e discussões

Nesta seção apresentamos uma leitura dos resultados, na abordagem clássica, e posteriormente na abordagem da TRI unidimensional. Foram utilizados os softwares BILOG-MG e pacote LTM do R, para dados dicotômicos em uma única população.

A primeira idéia na análise dos dados sob a perspectiva da abordagem clássica, foi avaliar a qualidade do instrumento utilizado, no que diz respeito à precisão do mesmo. O procedimento clássico com esta finalidade é a análise do coeficiente alfa de Cronbach, que mostrou valores do coeficiente variando de 0,8094 a 0,8176, indicando uma boa consistência interna do instrumento utilizado.

Primi (2012) defende a idéia de que, quando analisamos somente o índice de dificuldade -

ID de uma questão da prova, um bom item é aquele que apresenta alta variância. Os itens 84, 60, 31, 83, 1, 65, 86 e 40 apresentam variâncias abaixo de 0,15.

Um outro procedimento importante na análise exploratória de dados, diz respeito ao estudo da capacidade de discriminação do item em relação ao resultado do teste. Para este procedimento utiliza-se a correlação bisserial para verificar se uma determinada variável apresenta correlação significativa com o escore bruto produzido pelo conjunto dos itens.

Ao se analisar a correlação bisserial é possível observar os itens 11, 32, 52, 64, 85 e 87 com correlações negativas. Estes resultados permitem concluir que tais itens não apresentam boa discriminação para o índice que se pretende produzir, devendo ser excluídas de qualquer análise posterior.

Após a exclusão dos itens que pouco contribuíssem com a medida de proficiência, os itens resultantes apresentaram valores calibrados para os parâmetros apresentados na Tabela 1.

Tabela 1: Estimativa dos parâmetros dos 70 itens válidos. Vestibular ESALQ/USP 2012

ITEM	a	EP(a)	b	EP(b)	c	EP(c)	ITEM	a	EP(a)	b	EP(b)	c	EP(c)
I03	1,280	0,227	1,220	0,112	0,293	0,032	I45	1,381	0,171	0,505	0,107	0,233	0,038
I04	0,925	0,128	-0,172	0,279	0,328	0,072	I46	1,097	0,342	3,131	0,489	0,221	0,018
I05	1,463	0,248	1,303	0,092	0,250	0,026	I47	1,230	0,152	-0,192	0,198	0,335	0,062
I06	0,868	0,173	1,534	0,173	0,233	0,042	I48	1,493	0,181	-0,072	0,140	0,290	0,051
I08	0,905	0,151	0,396	0,252	0,323	0,063	I49	0,967	0,300	3,033	0,475	0,255	0,023
I09	1,192	0,199	1,641	0,118	0,185	0,026	I50	0,890	0,144	1,516	0,144	0,158	0,035
I10	1,375	0,195	0,477	0,127	0,307	0,042	I53	1,542	0,156	0,118	0,094	0,177	0,038
I12	1,215	0,208	1,437	0,117	0,255	0,029	I54	1,153	0,133	-0,236	0,190	0,268	0,062
I13	1,333	0,170	0,512	0,117	0,254	0,040	I55	1,414	0,145	-0,887	0,198	0,323	0,076
I15	0,754	0,148	1,272	0,225	0,259	0,053	I57	1,371	0,152	-0,323	0,157	0,259	0,058
I16	1,227	0,208	1,722	0,122	0,157	0,024	I58	0,664	0,171	2,290	0,314	0,273	0,046
I17	1,121	0,215	1,051	0,145	0,323	0,041	I59	2,012	0,730	2,917	0,364	0,243	0,011
I18	1,183	0,168	1,322	0,098	0,155	0,028	I60	1,728	0,405	2,701	0,265	0,073	0,009
I19	2,094	0,313	1,524	0,073	0,147	0,014	I61	1,394	0,340	2,373	0,237	0,205	0,017
I21	1,392	0,222	1,003	0,096	0,258	0,032	I63	1,208	0,261	2,458	0,239	0,131	0,018
I22	0,755	0,112	0,706	0,213	0,205	0,054	I65	0,867	0,309	4,109	0,924	0,201	0,018
I23	1,425	0,195	0,075	0,157	0,361	0,050	I67	0,772	0,253	3,772	0,726	0,232	0,024
I24	1,018	0,195	1,090	0,169	0,332	0,044	I68	1,083	0,340	2,740	0,406	0,282	0,023
I25	1,649	0,274	1,617	0,107	0,300	0,020	I69	1,272	0,440	3,347	0,579	0,207	0,014
I26	1,205	0,229	1,586	0,127	0,251	0,029	I70	1,497	0,204	0,695	0,095	0,263	0,033
I27	1,273	0,182	0,951	0,099	0,217	0,033	I72	0,851	0,147	0,467	0,265	0,316	0,064
I28	1,674	0,411	2,022	0,171	0,299	0,018	I73	1,292	0,146	-0,041	0,151	0,261	0,053
I29	1,465	0,305	2,048	0,161	0,205	0,018	I74	1,630	0,176	-0,099	0,116	0,253	0,046
I30	1,407	0,442	2,970	0,420	0,217	0,014	I75	1,449	0,209	1,363	0,085	0,181	0,023
I31	2,035	0,462	2,436	0,199	0,087	0,009	I76	1,333	0,166	0,297	0,130	0,265	0,045
I33	1,011	0,176	1,157	0,147	0,254	0,041	I78	1,309	0,141	-0,518	0,179	0,272	0,065
I35	0,963	0,126	0,409	0,180	0,230	0,052	I80	1,441	0,178	1,039	0,079	0,179	0,026
I36	1,037	0,249	2,218	0,230	0,225	0,027	I81	1,077	0,230	2,111	0,192	0,200	0,026
I38	1,055	0,174	1,092	0,134	0,236	0,040	I82	1,616	0,345	2,008	0,153	0,243	0,017
I39	1,369	0,229	0,668	0,138	0,400	0,039	I83	1,358	0,510	3,222	0,589	0,283	0,015
I40	0,749	0,284	5,324	1,535	0,161	0,014	I84	1,662	0,355	2,517	0,215	0,115	0,011
I41	0,844	0,124	0,667	0,199	0,223	0,053	I86	1,339	0,240	1,143	0,107	0,302	0,032
I42	1,100	0,173	0,474	0,175	0,294	0,052	I88	1,600	0,346	1,972	0,153	0,198	0,018
I43	1,435	0,182	0,531	0,101	0,236	0,037	I89	1,322	0,538	3,93	0,880	0,215	0,011
I44	1,218	0,435	3,446	0,646	0,216	0,014	I90	0,866	0,145	1,079	0,189	0,257	0,048

Uma outra medida útil no processo de avaliação do teste utilizado é a função de informação do teste $I(\theta)$. Para o exame seletivo vestibular aqui descrito, a curva de informação do teste é mostrada pelo traço contínuo na Figura 1, juntamente com a curva de erro padrão, mostrada em traços pontilhados.

Na Figura 1 os itens avaliados são mais informativos para os candidatos com escores entre 0

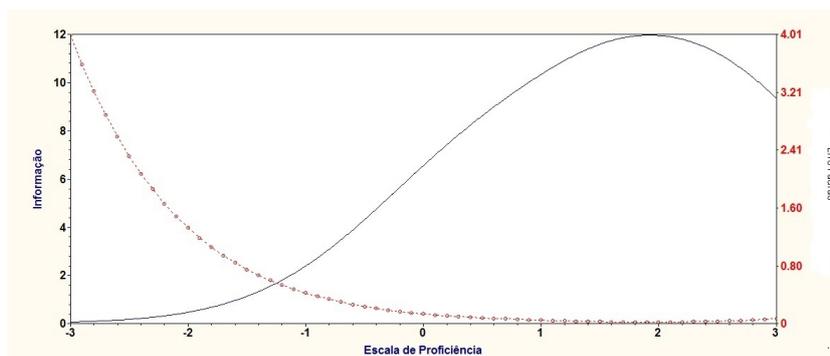


Figura 1: Curva de informação do teste

e 3. A forma assimétrica negativa da curva mostra que a prova foi bem ajustada para candidatos com maiores proficiências. Entretanto, na região esquerda da figura, a curva do erro padrão da estimativa é superior à curva de informação do teste, indicando no teste a ausência de itens que avaliem sujeitos com baixas proficiências.

Os valores estimados para os parâmetros dos itens viabilizaram obter as probabilidades de resposta correta, para diferentes níveis do traço latente, que puderam ser utilizadas para encontrar os itens âncora e quase âncora. Essa escala de proficiência, sintetiza o domínio dos conteúdos e habilidades alcançados pelos candidatos, o que permite inferir o nível de domínio das competências avaliadas.

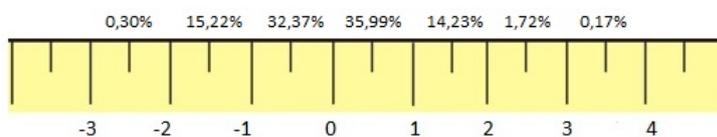


Figura 2: Níveis de proficiências e percentuais de participação dos avaliandos na escala.

Na prática, foram considerados somente quatro níveis (I, II, III e IV), correspondendo aos valores 0, 1, 2 e 3, respectivamente, da figura 2.1.

Uma proposta de interpretação pedagógica mais específica, utiliza os PCN's para o Ensino médio (2000), e é apresentada nos tópicos a seguir.

- **Nível I:** “Analisar, interpretar e aplicar os recursos expressivos das linguagens, relacionando textos com seus contextos, mediante a natureza, função, organização e estrutura das manifestações de acordo com as condições de produção e recepção.”;
- **Nível II:** Além da proficiência do Nível I, neste nível o candidato deve saber “Apropriar-se dos conhecimentos da Física, da Química e da Biologia, e aplicar esses conhecimentos para explicar o funcionamento do mundo natural, planejar, executar e avaliar ações de intervenção na realidade natural.”, além disso, deve também “Compreender o desenvolvimento da sociedade como processo de ocupação de espaços físicos e as relações da vida humana com a paisagem, em seus desdobramentos político-sociais, culturais, econômicos e humanos.”;

- **Nível III:** Além das proficiências dos Níveis I e II, neste nível o candidato deve “*Compreender a produção e o papel histórico das instituições sociais, políticas e econômicas, associando-as às práticas dos diferentes grupos e atores sociais, aos princípios que regulam a convivência em sociedade, aos direitos e deveres da cidadania, à justiça e à distribuição dos benefícios econômicos.*”;
- **Nível IV:** Além das proficiências dos Níveis I, II e III, neste nível o candidato deve “*Entender e aplicar métodos e procedimentos próprios das Ciências Naturais.*”.

Pode-se observar que as interpretações pedagógicas dos níveis de proficiência são traduzidas como um processo em que o conhecimento é acumulativo, o que é justificável, uma vez que o modelo logístico parte do princípio de que, quanto maior a proficiência do respondente, maior sua probabilidade de acerto. O traço latente do indivíduo representa a proficiência mínima que um respondente deve possuir para que sua probabilidade de acerto seja alta.

2.2 Conclusões

O uso da TRI mostra-se como uma alternativa interessante na análise de itens em processos seletivos, por considerar características importantes dos itens que não costumam ser contempladas em situações que se utilizam a Teoria Clássica dos Testes como critério de avaliação e classificação de candidatos. Mostra-se vantajosa como técnica que diagnostica itens importantes na mensuração do traço latente, identificando também aqueles que não o são. No caso do exame vestibular da FUVEST 2012 para candidatos aos cursos da ESALQ, identificou 70 dos 90 itens do teste, como importantes para construção de uma escala de proficiências, que poderia ser utilizada no processo de seleção daqueles candidatos.

Referências

- [1] Andrade, D.F., Tavares, H.R., Valle, R.C. (2000). Teoria da Resposta ao Item: Conceitos e Aplicações. *SINAPE*, 154p.
- [2] Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, **33**, 529-554
- [3] Primi, R. (2012). Psicometria: fundamentos matemáticos da teoria clássica dos testes. *Revista Avaliação Psicológica*, **11(2)**, 297-307